

XXV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXV ENANCIB

GT 8 – Dados, Informação e Tecnologia

FALSOS POSITIVOS NA DETECÇÃO DE TEXTOS GERADOS POR IA: ANÁLISE DO ENANCIB 2019

FALSE POSITIVES IN THE DETECTION OF AI-GENERATED TEXTS: ANALYSIS OF ENANCIB 2019

Isaías Braga – Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)/UFRJ

Fábio Gouveia – Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)

Tiago Braga – Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)

Modalidade: Resumo Expandido

Resumo: o uso de Inteligência Artificial Generativa na produção de trabalhos acadêmicos não é uma prática esperada para comunicações científicas. Avaliar softwares de Detecção de Uso de IA Generativa em textos acadêmicos é fundamental para o uso desse tipo de aplicação por professores e avaliadores. Essa pesquisa analisa a ocorrência de “Falsos positivos” na detecção de uso de IA Generativa na produção de textos no contexto acadêmico. Ela empregou uma análise quali-quantitativa para a avaliação de 341 resumos de trabalhos do XXI ENANCIB. Os softwares apresentaram variações relevantes na atribuição de seus resultados, embora a maioria indique valores elevados de acerto.

Palavras-chave: inteligência artificial generativa; detecção de IA em textos acadêmicos; falsos positivos.

Abstract: the use of Generative Artificial Intelligence in the production of academic papers is not an expected practice for scientific communications. Evaluating software for detecting the use of Generative AI in academic texts is essential for the use of this type of application by teachers and evaluators. This research analyzes the occurrence of “False positives” in the detection of the use of Generative AI in the production of texts in the academic context. It used a quali-quantitative analysis to evaluate 341 abstracts of papers from the XXI ENANCIB. The software presented relevant variations in the attribution of their results, although most indicate high accuracy values.

Keywords: generative artificial intelligence; AI detection in academic texts; false positives.

1 INTRODUÇÃO

Desde a publicação do *preprint* “*Attention is all you need*” (Vaswani *et al.*, 2017), elaborado por pesquisadores do Google, houve progressivamente uma popularização das ferramentas de IA Generativa, ocasionando uma mudança no modo como alunos e pesquisadores produzem seus trabalhos acadêmicos. A efetividade dessas ferramentas de gerar textos originais, com características similares aos textos gerados por humanos, pode ser utilizada para a produção de parte ou todo o texto de comunicação de pesquisas científicas.

Por mais que a responsabilidade desses trabalhos seja ainda assim corretamente vinculada ao autor que apresenta o texto (Braga, I.; Braga, T., 2024) e que possa cumprir o objetivo de avançar a conversa científica do campo, é interessante notar que aspectos de aprendizagem e desenvolvimento da habilidade de escrita e do raciocínio complexo envolvido nesse tipo de trabalho pode ser subvertido com o uso das ferramentas de IA Generativa.

Ainda assim, essa prática não é a esperada para comunicações científicas, o que tem gerado orientações sobre o uso aceitável de IA Generativa por diversos editores científicos, resultando em políticas que normatizam o seu uso na construção de produções acadêmicas. Empresas fornecem software com o objetivo de verificar se os textos submetidos foram ou não produzidos por ferramentas de IA Generativa, enquanto as próprias ferramentas de IA que produzem textos também podem ser utilizadas para detectar se algum texto foi ou não gerado com o seu auxílio ou de alguma ferramenta similar.

Entretanto, não há certeza se o uso das ferramentas de IA Generativa para a detecção do seu uso não poderiam gerar resultados de “Falso positivo”, classificando textos gerados por humanos como tendo sido produzidos por uma máquina. Esses resultados podem induzir os usuários de softwares que objetivam à detecção de uso de IA Generativa a duvidarem dos dados produzidos, tanto quando marcados como positivo, quanto, em situação ainda pior, quando o uso de IA Generativa é erroneamente atribuído, prejudicando os autores desses textos e suas carreiras acadêmicas.

Avaliar a eficiência de Detecção de Uso de IA Generativa (DUIAG) em textos acadêmicos é uma etapa fundamental para aumentar a confiabilidade no uso dessas ferramentas. Para evitar um conflito de interesse de empresas produzindo trabalhos que lhes beneficiem economicamente ao demonstrar a efetividade de suas próprias ferramentas, análises independentes devem ser realizadas. Outra justificativa para a realização de testes independentes é derivada do fato de que a maior parte das ferramentas de IA Generativa são majoritariamente comerciais e estrangeiras e treinam seus algoritmos de IA a partir de um corpus que não possui transparência dos trabalhos utilizados, o que levanta uma questão adicional, se possuem de fato uma capacidade adequada de avaliar textos de outros idiomas que não os textos em inglês ou da língua nativa do país onde o software foi desenvolvido.

O objetivo dessa pesquisa é analisar a ocorrência de “Falsos positivos” na detecção de uso de IA Generativa na produção de textos no contexto acadêmico. Para tal, selecionamos

um conjunto de artigos que, pela época em que foram gerados, os autores não podem ter feito uso de ferramentas do tipo para auxiliar na sua produção.

2 REVISÃO DE LITERATURA

2.1 Inteligência artificial

O campo de “Inteligência Artificial” não possui uma definição clara e amplamente aceita. Desde o batismo desse campo de estudo, em uma conferência em 1956, não foi estabelecido uma definição clara e objetiva do que se tratava a área, seus objetivos e atuação (Scaruffi, 2018). Entretanto, o termo passou a designar “todas as técnicas que permitem que uma máquina simule a inteligência humana, especialmente para aprender, prever, tomar decisões e perceber o mundo ao redor” (Abrassart *et al.*, 2018, p. 20), ou “a habilidade de fazer coisas que requeiram inteligência quando feitas por humanos” (Davies, 2011, p. 603). Segundo Zhang *et al.* (2023, p. 6), a essência da inteligência artificial está na “busca para criar máquinas que consigam imitar, replicar ou até superar as habilidades cognitivas humanas”.

É no contexto de máquinas executarem tarefas que antes eram exclusivas de cérebros humanos que surge a IA Generativa. Alicerçada em diversos avanços interdisciplinares, a IA Generativa surpreendeu a sociedade ao criar documentos originais, muitas vezes indistinguíveis de documentos criados por humanos. Estas IA's são treinadas e alimentadas com enormes massas de informação e desenvolvem grandes modelos de linguagem (*Large Language Model* - LLM) para gerar resultados antes inimagináveis, tanto pela verossimilhança com documentos criados por humanos, quanto pela qualidade dos resultados produzidos.

Os grandes modelos de linguagem (Large Language Model – LLM) são a evolução de outros modelos criados para compreender a linguagem humana natural. Naveed *et al.* (2024) afirma que os LLMs são uma evolução dos modelos de linguagem pré-treinados, que por sua vez evoluíram dos modelos neurais de linguagem e dos modelos estatísticos. De forma resumida os LLMs são algoritmos especializados em compreender grandes quantidades de textos, entendendo a relação e importância dos conceitos presentes nos documentos. São pré-treinados analisando bases de dados gigantescas e, atualmente, envolve o ajuste de trilhões de parâmetros, com o objetivo de compreender entradas, em linguagem natural, feita pelos usuários e produzir resultados que replicam os padrões aprendidos por meio da base de dados inicialmente utilizadas para o treinamento.

Uma série de algoritmos desse tipo, batizada de GPT, foram lançados pela Open AI a partir de 2018, o primeiro deles, ChatGPT 1.0 teve seu treinamento alicerçado em uma base de dados com 117 milhões de parâmetros. Sua segunda versão, lançada já em 2019, contava com uma base de dados com 1,5 bilhões de parâmetros, e sua terceira versão, lançada em 2022 (a primeira aberta ao público em geral), com 175 bilhões de parâmetros. Na sua quarta versão, de 2024, já utilizava uma base de dados com 1,76 trilhões de parâmetros (Zhang *et al.*, 2023), demonstrando um aumento vertiginoso no tamanho das bases de dados utilizadas.

Os sistemas de Inteligência Artificial Generativa se popularizaram desde sua introdução ao público, no final de 2022 (Gao *et al.*, 2023). O lançamento da ferramenta ChatGPT pode ser considerada um marco para a disseminação dessa tecnologia para uma grande massa de usuários ao redor do mundo. Esse tipo de tecnologia se concentra na criação de novos conteúdos a partir do aprendizado de padrões em dados já existentes.

O treinamento de algoritmos como o do ChatGPT faz uso de grandes volumes de conteúdo, especialmente de texto não classificado, para que o código seja treinado ao encontrar padrões e relações entre esses textos e então produza os Large Language Model¹ (LLM). As fontes de dados utilizadas para esse treinamento são fundamentais para o aprendizado do algoritmo, pois a qualidade dessas informações trará impacto no desempenho do modelo representacional da inteligência artificial (Gao *et al.*, 2023). Diversas fontes de informação são utilizadas no treinamento dos algoritmos de inteligência artificial e construção do LLM, tanto públicas quanto privadas. Alguns exemplos são a Wikipédia, livros, artigos acadêmicos, notícias, posts de mídia social etc. (Braga, 2024).

2.2 Detecção do uso de IA Generativa

Diversos autores têm expressado a preocupação de alunos e pesquisadores estarem utilizando ferramentas de IA Generativa para a elaboração de seus trabalhos. Walters (2023) faz uma revisão de diversos trabalhos e executa uma pesquisa analisando especificamente a eficácia de softwares específicos para DUIAG, apontando três softwares com os melhores resultados na identificação e rotulação de textos acadêmicos quanto ao uso de IA Generativa.

¹ LLM são um componente de ferramentas de IA generativa. São projetados para gerar texto. Eles são treinados em enormes conjuntos de dados de texto, a fim de aprender padrões e estruturas na linguagem. Uma vez treinados, eles podem gerar novos textos que são coerentes e seguindo o contexto dado. Disponível em <https://blog.dsacademy.com.br/llms-e-a-evolucao-da-ia-generativa/>. Acesso em: 28 jan. 2025.

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

Embora a pesquisa de Walters (2023) identifique os melhores softwares, ela não aborda três aspectos importantes: não abrange textos de outros idiomas, não analisa os “falsos positivos” e não emprega softwares de IA Generativa gerais ou gratuitos para a análise, exigindo ferramentas pagas e ou inacessíveis para todos os professores e avaliadores. Entende-se que é necessário avançar nesses pontos, a fim de permitir que a identificação de textos criados por IA Generativa possa ser realizada de forma mais abrangente.

3 METODOLOGIA

A elaboração dessa pesquisa envolveu dois critérios principais para a seleção de trabalhos, visando a identificação de ocorrências de “falsos positivos” em DUIAG: textos seguramente produzidos por humanos e textos em português.

Assim, a metodologia dessa pesquisa envolveu a avaliação de 341 resumos de trabalhos apresentados no XXI ENANCIB, ocorrido entre 25 e 29 de outubro de 2021. Portanto, antes do lançamento do ChatGPT ao público geral, a primeira ferramenta de IA Generativa adotada para criação de textos, em novembro de 2022. O Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) é tido como o principal evento de pesquisa e de pós-graduação da área de Ciência da Informação no Brasil, tendo quase exclusivamente trabalhos em português em seus anais.

Em relação às ferramentas de IA Generativa utilizadas para a DUIAG, foram selecionadas quatro ferramentas: 1) ChatGPT, da OpenAI, por ser a ferramenta pioneira de acesso ao público geral e por sua popularidade entre alunos e professores; 2) Gemini, do Google, por estar disponível para muitos professores via parcerias da empresa com universidades; 3) DeepSeek, da DeepSeek, por ser uma das mais populares ferramenta de código aberto e por ser totalmente gratuita aos usuários; e 4) DuckDuckGo, da DuckDuckGo, por ser amplamente utilizado e também gratuito, oferecendo acesso à uma versão distinta de LLM da OpenAI. Em todas as ferramentas, foi mantido o modelo LLM padrão para as análises: GPT-4-turbo (gpt-4.5), Gemini 2.0 Flash, DeepSeek-V3 e GPT4o mini.

O uso dessas ferramentas para a avaliação dos textos ocorreu com o uso do seguinte prompt e subsequente envio do texto a ser analisado: “Analisar individualmente cada texto a seguir e informe qual é a probabilidade de ter sido escrito integralmente por humanos. Informe essa probabilidade atribuindo a cada texto um número de 0% a 100%”. Os resultados foram tabulados e analisados estatisticamente pelos autores.

Após o prompt inicial, foi utilizada uma variação do prompt com vistas a identificar os critérios utilizados para atribuição da probabilidade pelo software. O segundo prompt utilizado nas quatro ferramentas foi: “Irei passar uma lista de textos para que você analise individualmente cada um deles e gere uma listagem simples, apenas com o número do texto e a probabilidade, informando qual é a probabilidade de cada um deles ter sido escrito integralmente por humanos. Informe essa probabilidade atribuindo a cada texto um número de 0% a 100%. Como e quais critérios você utilizará para fazer a análise dessas probabilidades?”.

Por fim, visando validar as respostas, os 341 resumos foram submetidos a uma plataforma específica de identificação de textos criados por IA Generativa, chamada zeroGPT.

3 RESULTADOS E DISCUSSÃO

Ao todo, foram analisados 341 resumos de trabalhos. Os resumos possuíam média de 1.133 caracteres, em 164 palavras. Separados em quartis, os resultados da análise de probabilidade para a Detecção de Utilização de IA Generativa nos textos foi a seguinte.

Quadro 1 - Probabilidade do texto ser escrito por humanos: análise em quartis

Probabilidade	ChatGPT	Gemini	DeepSeek	DuckDuckGo
0% a 25%	0	0	0	0
26% a 50%	1	0	0	0
51% a 75%	32	0	17	0
76% a 100%	308	341	324	341

Fonte: elaborado pelos autores (2025)

Como a maior parte dos resultados se concentrou no último percentil, assim, foi realizada uma análise mais granular com foco nessa secção.

Quadro 2 - Probabilidade do texto ser escrito por humanos: análise granular

Probabilidade	ChatGPT	Gemini	DeepSeek	DuckDuckGo
<70%	8	0	4	0
70% a 79%	26	0	21	0
80% a 89%	171	9	137	189
90% a 100%	136	332	179	152

Fonte: elaborado pelos autores (2025)

Os valores de probabilidade variaram da seguinte forma para cada ferramenta, sendo os menores e maiores valores atribuídos a um texto:

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

- ChatGPT: 50% e 95%;
- Gemini: 80% e 98%;
- DeepSeek: 55% e 99%;
- DuckDuckGo: 80% e 95%.

Apesar das ferramentas de IA Generativa utilizadas para a DUIAG não terem sido desenvolvidas especificamente para a tarefa abordada, podem facilmente ser empregadas por professores e pesquisadores para realizar tal atividade, uma vez que estão amplamente disponíveis para esses profissionais. Foi possível questionar diretamente as próprias ferramentas para entender como executaram o processo de análise e atribuição da probabilidade do texto ter sido gerado por humanos.

Ao ser aplicado o segundo prompt da pesquisa, foram sinalizados pelas quatro ferramentas analisadas os critérios a seguir:

- ChatGPT (Originalidade e variação lexical; Coesão e coerência natural; Erros e imperfeições humanas; Uso de referências, fontes e citações; Estilo e subjetividade e Padrões típicos de IA);
- Gemini (Complexidade e Naturalidade da Linguagem – Vocabulário, Estrutura das Frases, Fluidez e Coerência, Expressões Idiomáticas e Culturais; Estilo e Tom - Consistência do Estilo, Expressão Emocional e Subjetividade e Humor e Ironia; Informação e Contexto - Conhecimento de Mundo e Atualidade, Originalidade e Novidade e Erros e Inconsistências Naturais);
- DeepSeek (Fluidez e Coerência; Originalidade e Criatividade; Contexto e Profundidade Emocional; Erros e Imperfeições; Uso de Linguagem Técnica ou Especializada e Marcadores de IA (Repetição, Estrutura Rígida, etc.));
- DuckDuckGo (Coerência e Coesão; Complexidade Linguística; Erros e Estilo; Originalidade e Contexto e Relevância).

Quadro 3 - Critérios informados pelas ferramentas de IA Generativa para DUIAG

Critérios	ChatGPT	Gemini	DeepSeek	DuckDuckGo
Estilo	- Coesão e coerência natural - Estilo e subjetividade	- Estrutura das Frases - Fluidez e Coerência - Expressões Idiomáticas e Culturais - Consistência do Estilo - Expressão Emocional	- Fluidez e Coerência - Uso de Linguagem Técnica ou Especializada	- Coerência e Coesão -Complexidade Linguística

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

		e Subjetividade - Humor e Ironia - Vocabulário		
Originalidade	- Originalidade e variação lexical	- Originalidade e Novidade	- Originalidade e Criatividade	- Originalidade
Contexto	- Uso de referências, fontes e citações	- Conhecimento de Mundo e Atualidade	- Contexto e Profundidade Emocional	- Contexto e Relevância
Marcadores Humanos	- Erros e imperfeições humanas	- Erros e Inconsistências Naturais	- Erros e Imperfeições	- Erros e Estilo
Marcadores de IA	- Padrões típicos de IA		- Marcadores de IA (Repetição, Estrutura Rígida, etc.)	

Fonte: elaborado pelos autores (2025)

Ao se analisar os critérios utilizados e o resultado da confirmação via novo prompt, percebe-se que os critérios têm natureza qualitativa.

Subsequente às análises feitas nas 4 ferramentas de IA Generativa, para fins de comparação, todos os textos foram avaliados novamente em uma ferramenta específica para análise de textos para a DUIAG, chamada ZeroGPT, que oferece recursos gratuitos de forma limitada. Como os textos consistiam apenas do resumo dos artigos analisados, o limite de caracteres imposto pela ferramenta não foi um impeditivo para a análise proposta neste estudo. Segundo a empresa responsável pelo zeroGPT, o algoritmo de detecção deles foi desenvolvido com a análise de mais de 10 milhões de artigos e textos e possui uma precisão de 98% na identificação correta de textos, aplicável a todos os idiomas². Apesar de ser um software desenvolvido especificamente para a tarefa, foi o que apresentou os piores resultados, uma vez que todos os textos da amostra selecionada para análise foram integralmente produzidos por humanos.

Quadro 4 - Probabilidade do texto ser escrito por humanos: ZeroGPT

Probabilidade	ZeroGPT
0% a 25%	54
26% a 50%	30

² Disponível em: <https://www.zerogpt.com>

51% a 75%	30
76% a 100%	227

Fonte: elaborado pelos autores (2025)

Apesar de ter apontado que 39 dos 54 textos do quartil de 0% a 25% teriam 0% de probabilidade de terem sido produzidos por humanos, consequentemente tendo sido elaborados exclusivamente por IA Generativa, foi a única ferramenta que apontou textos com 100% de certeza de terem sido escritos inteiramente por humanos, ao todo, 219 resumos receberam esta probabilidade dentre os 227 do quartil de 76% a 100%.

4 CONCLUSÕES

O uso de softwares de IA Generativa para a produção de textos acadêmicos tem se tornado mais usual por estudantes e pesquisadores e ferramentas para identificar o uso de tais tecnologias têm sido cogitadas e empregadas por professores e avaliadores em diferentes contextos acadêmicos. Entretanto, o uso de ferramentas para detectar o emprego de softwares de IA Generativa para a elaboração de textos trás sérias incertezas relacionadas à capacidade desses softwares em detectar a utilização do uso de ferramentas de IA Generativa na produção dos textos. Uma das principais incertezas diz respeito à ocorrência de “falsos positivos”, que podem ter um efeito devastador e com impactos imediatos na vida acadêmica do aluno ou pesquisador, como a recusa do trabalho, bem como com efeitos de médio e longo prazo em suas carreiras profissionais e acadêmicas.

Na conclusão desse trabalho, identifica-se que o uso de softwares de IA Generativa de uso geral para a Detecção de Uso de IA Generativa (DUIAG) em textos acadêmicos não é infalível. Os softwares apresentaram variações relevantes na atribuição de suas probabilidades, ainda que a maioria deles indique valores elevados de acerto. As oscilações ocorrem mesmo analisando textos que, conforme comprovado, foram exclusivamente escritos por humanos, de acordo com a condição garantida pelo recorte temporal detalhado na metodologia desta pesquisa para seleção do corpus textual. Os resultados foram ainda piores ao utilizar um software específico para a DUIAG, quando 11% dos textos analisados foram rotulados como sendo produzidos integralmente por IA Generativa.

Assim como qualquer trabalho acadêmico, esta pesquisa apresenta limitações que devem ser cuidadosamente consideradas para uma interpretação adequada dos resultados. Dentre as limitações mais relevantes, destaca-se o uso de softwares de inteligência artificial

generativa (IA Generativa) de propósito geral para a Detecção de Uso Indevido de Inteligência Artificial Generativa (DUIAG) nos textos analisados. Essas ferramentas, por não serem projetadas especificamente para essa finalidade, enfrentam restrições significativas em sua capacidade de realizar uma análise aprofundada e precisa. Em particular, a análise qualitativa dos textos revelou-se insuficiente, com falhas na identificação de nuances contextuais, estilos de escrita ou padrões específicos que poderiam indicar o uso de IA Generativa de forma mais confiável.

Um dos principais motivos para a ocorrência de falsos positivos está relacionado à generalidade dessas ferramentas. Softwares de IA Generativa de propósito geral, como os utilizados na pesquisa, são otimizados para tarefas amplas, como geração de texto ou resposta a perguntas, e não para a detecção específica de conteúdo gerado por IA. Isso resulta em algoritmos que podem interpretar erroneamente características textuais comuns, como padrões sintáticos ou vocabulário técnico frequente em textos acadêmicos, como indicadores de uso de IA, mesmo quando o texto foi produzido por humanos. Essa limitação é agravada pela falta de estratégias refinadas de análise, como a capacidade de diferenciar entre textos humanos e gerados por IA com base em critérios mais sofisticados, como a coerência discursiva ou a intencionalidade autoral.

Além disso, mesmo o uso de um software especializado em DUIAG não conseguiu superar essas limitações, sugerindo que as tecnologias atuais ainda enfrentam desafios intrínsecos na distinção precisa entre textos humanos e gerados por IA. Esse desempenho aquém do esperado pode ser atribuído a fatores como a complexidade inerente à detecção de IA Generativa, especialmente em textos acadêmicos, que frequentemente apresentam um alto grau de formalidade, estruturação e vocabulário técnico. Esses elementos podem confundir os algoritmos, levando a falsos positivos, já que os modelos podem interpretar características legítimas de textos acadêmicos como sinais de geração artificial.

Outro fator determinante para a ocorrência de falsos positivos foi a escolha de analisar apenas os resumos dos trabalhos acadêmicos, que são textos curtos por natureza. A brevidade desses textos pode limitar a quantidade de informações contextuais disponíveis para as ferramentas de análise, reduzindo sua capacidade de identificar padrões robustos que diferenciem textos humanos de textos gerados por IA. Ferramentas de DUIAG tendem a apresentar melhor desempenho em textos mais longos, que contêm maior densidade de elementos linguísticos, como variações estilísticas, argumentos encadeados ou marcadores

de autoria, permitindo uma avaliação mais abrangente. Em resumos, a escassez de conteúdo pode levar a uma análise superficial, aumentando a probabilidade de erros de classificação, como falsos positivos, especialmente quando os textos contêm vocabulário técnico ou estruturas padronizadas comuns em contextos acadêmicos.

Além disso, a qualidade e a diversidade do conjunto de dados utilizado para treinar os modelos de IA também desempenham um papel crucial. Modelos treinados em conjuntos de dados limitados ou não representativos podem apresentar viés, identificando erroneamente textos humanos como gerados por IA devido à falta de familiaridade com certos estilos de escrita ou gêneros textuais. Por exemplo, resumos acadêmicos frequentemente seguem convenções rígidas, com linguagem objetiva e padronizada, que podem se assemelhar a padrões encontrados em textos gerados por IA, contribuindo para a ocorrência de falsos positivos.

Por fim, é importante considerar que as limitações observadas não se restringem apenas às ferramentas utilizadas, mas também ao próprio estado atual da tecnologia de detecção de IA Generativa. A rápida evolução das capacidades de geração de texto por IA torna a tarefa de detecção cada vez mais desafiadora, exigindo avanços contínuos em algoritmos e metodologias. Assim, os resultados desta pesquisa refletem não apenas as limitações das ferramentas empregadas, mas também a complexidade inerente à tarefa de distinguir textos humanos de textos gerados por IA em contextos acadêmicos, especialmente quando se trabalha com amostras textuais curtas e altamente estruturadas.

Considerando essas limitações, sugere-se que a continuidade desse trabalho possa envolver o uso de outros softwares específicos para DUIAG, especialmente ferramentas pagas. Também se acredita que o uso de trabalhos completos para a análise possa fornecer resultados melhores em relação à rotulação do uso de IA Generativa em sua elaboração.

Por fim, entende-se que a utilização de IA Generativa para produção e melhoria de textos é uma tendência na sociedade, como apontam Braga e Braga (2024). Nesse sentido, é importante compreender que mesmo a identificação de probabilidades altas de textos criados por IA Generativa não condicionam a que tais textos tenham, necessariamente, sido concebidos por esse tipo de ferramenta, pois elas podem ser utilizadas como um instrumento de melhoria a partir de textos fornecidos por humanos.

REFERÊNCIAS

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

ABRASSART, C. *et al.* **Declaração de Montreal pela IA responsável.** [S. l.: s. n.], 2018. Disponível em: https://declarationmontreal-iaresponsable.com/wp-content/uploads/2023/01/Portugues_-_UdeM_Decl_IA_Resp_LA_Declaration_vf.pdf. Acesso em: 11 jul. 2024.

BRAGA, I. **Revisão sistemática dos modelos de inteligência artificial generativa LLM.** [s. l.], 2024.

BRAGA, I.; BRAGA, T. Inteligência Artificial Generativa: uma discussão sobre autoria. *In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO*, 24., 2024, Vitória, ES. **Anais [...]**. Vitória, ES: ANCIB, 2024. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxivenancib/paper/view/2605>. Acesso em: 24 fev. 2025.

DAVIES, C. R. An evolutionary step in intellectual property rights – Artificial intelligence and intellectual property. **Computer Law & Security Review**, [s. l.], v. 27, n. 6, p. 601–619, 2011.

GAO, K. *et al.* **Examining User-Friendly and Open-Sourced Large GPT Models: A Survey on Language, Multimodal, and Scientific GPT Models.** [S. l.]: arXiv, 2023. Disponível em: <http://arxiv.org/abs/2308.14149>. Acesso em: 28 jun. 2024.

NAVEED, H. *et al.* **A Comprehensive Overview of Large Language Models.** [S. l.]: arXiv, 2024. Disponível em: <http://arxiv.org/abs/2307.06435>. Acesso em: 26 jun. 2024.

SCARUFFI, P. **Intelligence is not Artificial - Expanded Edition: a history of artificial intelligence and why the singularity is not coming any time soon.** Expanded edition. [S. l.]: CreateSpace Independent Publishing Platform, 2018.

VASWANI, A. *et al.* **Attention Is All You Need.** [S. l.]: arXiv, 2017. Disponível em: <https://arxiv.org/abs/1706.03762>. Acesso em: 24 maio 2025.

WALTERS, W. H. The Effectiveness of Software Designed to Detect AI-Generated Writing: A Comparison of 16 AI Text Detectors. **Open Information Science**, [s. l.], v. 7, n. 1, p. 20220158, 2023.

ZHANG, E. Y. *et al.* From Turing to Transformers: A Comprehensive Review and Tutorial on the Evolution and Applications of Generative Transformer Models. **Sci**, [s. l.], v. 5, n. 4, p. 46, 2023.